



TITLE:

毒性の重症度を考慮したがん第I相試験の用量探索方法の提案 (統計的モデルの新たな展望とそれに関連する話題)

AUTHOR(S):

岩田, 知子; 高橋, 史朗; 井上, 永介; 竹内, 正弘

CITATION:

岩田, 知子 ...[et al]. 毒性の重症度を考慮したがん第I相試験の用量探索方法の提案 (統計的モデルの新たな展望とそれに関連する話題). 数理解析研究所講究録 2012, 1804: 70-89

ISSUE DATE:

2012-08

URL:

<http://hdl.handle.net/2433/194382>

RIGHT:

毒性の重症度を考慮したがん第 I 相試験の用量探索方法の提案

北里大学大学院薬学研究科 臨床医学（臨床統計学）

Department of Clinical Medicine (Biostatistics),
Graduate School of Pharmaceutical Sciences, Kitasato University

岩田 知子 (Tomoko Iwata)

高橋 史朗 (Fumiaki Takahashi)

井上 永介 (Eisuke Inoue)

竹内 正弘 (Masahiro Takeuchi)

1. 序論

がん領域における第 I 相試験の目的は、新規薬剤の最大耐用量 (MTD: maximum tolerated dose) を決定することである。MTD 決定のためのエンドポイントは、用量制限毒性 (DLT: dose limiting toxicity) の発現の有無である。通常、NCI CTC の重症度分類でグレード 3 以上の DLT 発現割合が、事前に定めた割合以下になる最大の用量を最大耐用量とする。事前に定める割合とは、許容できる DLT 発現率の目標値であり、25~33%程度に設定されている。

伝統的な最大耐用量探索方法は 3+3 デザインである。この方法の利点として、用量増加ルールが簡便であり、MTD を超える投与量で処置される被験者数が少ないことがあげられる。欠点としては、毒性は発現しないが、効果が期待できないような低用量で、多くの被験者が処置されることがあげられる。試験を最大限安全に行うためにも、効果が期待できない用量で治療される患者の数は最小限に留められるべきである。この欠点を克服し、さらに MTD 推定精度を改善した数多くの方法が提唱されてきた。CRM (Continual Reassessment Method) は第 I 相試験デザインを改善するもう 1 つのアプローチである、ベイズ流デザインを用いた方法である。ベイズ流デザインは、複雑化する様々な臨床的要求に対して柔軟に対応でき、事前情報を積極的に活用することができる。この方法の利点は、それまでのすべての患者の治療の結果を、次の患者の用量を決定するのに用いることができるという点と、MTD の推定精度が改善できるという点にある。しかしながら、CRM デザインでは患者の不均一性が高い時でも、たった 1 人の患者の治療結果を基にして用量レベルをいくつか上げてしまうことがありうる、といった問題点がある。この問題点に着目

して基本的な CRM デザインの変法が提案されてきた。このように様々な方法が提案される一方で、Ji et al (2007)では、2006 年 4 月から 5 月の間に M.D. Anderson Cancer Center で実施された第 I 相試験 22 試験について用量探索方法を調査したところ、9 割を超える 20 試験で伝統的な 3+3 デザインが用いられている状況を報告した。その主たる理由として、複雑な統計モデルを必要としない簡便さをあげている。そして彼らは 3+3 デザインと同様に非常に簡便なベイズ流デザインを提案し、CRM などの方法と同程度の MTD 推定精度があることをシミュレーションで示した。

NCI CTC を用いてグレード 3 以上を一律に制限毒性とする判断方法の問題点として、次の二点が考えられる。一点目は、腎毒性や神経毒性ではグレード 3 とグレード 4 で大きく症状および重症度が異なるにもかかわらず、それらを同等に扱っている点である。二点目は、異なる二つの有害事象が同じグレードであっても、医師が感じる症状の重症度が異なる場合があるにもかかわらず、それらを同等に扱っていることである。例えば、高頻度で発現するが、危険度としてはあまり高くないような毒性である骨髄抑制のグレード 3 と、発現頻度は低い、危険度が高い毒性である肝毒性のグレード 3 では重症度が医師にとって 3 倍ほども違いがある。

そこで、本研究では、これらの重症度の問題点を解決するために、Ji らが提案した方法を拡張した毒性の重症度を考慮した方法を提案し、MTD の推定精度の観点から Ji らの方法と比較検討することを目的とした。

第 2 章では本研究の提案方法について述べる。第 3 章では、提案法と Ji らの方法をシミュレーションにより比較する。第 4 章ではシミュレーションの結果を示す。そして最終的な考察を第 5 章で行う。

<既存の代表的な用量探索デザイン>

➤ 3+3 デザイン

3 人中 DLT 発現数

- | | |
|---|---------------------------------------|
| { | 0 人 → 増量 |
| | 1 人 → もう 3 人に追加投与 |
| | DLT 発現 0 例 → 増量 |
| | 1 例以上 → MTD を超えたとして試験終了 (1 レベル下を推奨用量) |
| | 2, 3 人 → MTD を超えたとして終了 (1 レベル下を推奨用量) |

まず 3 人の患者が最初の投与レベルに登録される。もし最初の 3 人のうち 1 人も DLT が発現しなければ、次の用量レベルに 3 人が登録される。もし、1 人に DLT が発現すれば、同

じ用量レベルにあと 3 人登録される。2 人もしくは 3 人に DLT が発現すれば、その用量は MTD を上回ったと判断して、増量はそこで中止される。6 人が治療された場合、1 人にだけ DLT が発現すれば、増量は続けられる。そうでなければ、その用量は MTD を上回ったとして増量はそこで中止される。このプロセスを、MTD が決定するか、MTD を上回ると判定される最初の用量まで繰り返す。

➤ CRM (Continual Reassessment Method)

海外の試験結果などの事前情報を活用し、用量と制限毒性の発現確率の関係をモデル化したデザインである。以下に示すような、用量—反応曲線を仮定し、被験者の治療結果が得られる度に、毒性発現の有無から用量反応曲線を更新し、毒性発現確率をベイズ推定し、用量決定を行う。その推定 MTD に最も近いレベルで次の患者を治療する。前もって設定された数の患者の治療が終わった後、許容される DLT 発現率を超えない最大の用量を求める方法である。(O'Quigley, Pepe, Fisher, 1990)

用量反応モデルは次のような 1 パラメータのモデルで表わすことができる。

$$\Pr(Y_i = 1) = \psi(x_i, a)$$

$$\left\{ \begin{array}{ll} Y_i : \text{被験者 } i \text{ の制限毒性発現の有無} & \text{つまり、毒性発現あり } Y_i = 1 \\ & \text{発現なし } Y_i = 0 \\ x_i : \text{被験者 } i \text{ の治療に用いた用量} \\ a : \text{用量反応曲線を示すモデルパラメータ} \end{array} \right.$$

このモデルの設定条件は、用量が上がるにつれ、毒性も増加することである。 $\psi(x_i, a)$ は用量が上がるのに伴い単調増加することを前提としている。そのような用量—反応曲線として次のようなものがある。

$$1) \quad \psi(x_i, a) = \alpha_i^a \quad (i = 1, \dots, k) \quad 0 < \alpha_1 < \dots < \alpha_k < 1 \quad 0 < a < \infty$$

$$\alpha_i = (\tanh x_i + 1)/2 \quad (i = 1, \dots, k)$$

2) ロジスティックモデル

$$\psi(x_i, a) = \frac{\exp(a_0 + ax_i)}{1 + \exp(a_0 + ax_i)} \quad a_0 \text{ は定数}$$

2. 毒性の重症度を考慮した用量探索方法

以下、用量 ($i = 1, \dots, d$) における毒性発現確率を p_i 、目標値の毒性発現確率を p_t とおき、毒性の重症度を考慮した用量探索方法を提案法と呼ぶ。

2.1 Ji らの提案方法

本研究のもととなる先行研究である Ji らが提案した方法について説明する。Ji らの方法は、ベイズ流デザインであり、事前分布 $\beta(0.005, 0.005)$ を用い、 p_i の事後分布を計算する。得られる観察データのうち、用量 i で治療される患者 n_i 人、グレード 3 以上の毒性発現者を x_i 人とする、尤度関数は二項分布の密度関数に従い、次のように表わされる。

$$l(p) \propto \prod_{i=1}^d p_i^{x_i} (1 - p_i)^{n_i - x_i}$$

ここで、ある確率分布 $p(\mathbf{p} | \text{data})$ について、事後分布がその事前分布と同じ関数形になるような事前分布が存在するとき、この分布を共役事前分布という。ベータ分布は二項分布の共役事前分布であるため、事後分布は $B(0.005 + x_i, 0.005 + n_i - x_i)$ に従う。

その時点の用量での推定毒性発現確率によって、次の用量へ増量するか、同じ用量を維持するか、1つ下の用量へ減量するかを決定する。Ji らの方法の特徴的な点は毒性発現確率 $(0, 1)$ を次のような 3 区間 (Δ) に分けて、それぞれの区間の事後確率を計算し、用量増減を判断する点である。ここで、 σ_i を p_i の事後標準偏差、 k_1, k_2 は正の定数とする。

$$\Delta = \{(0, p_i - k_1 \sigma_i), [p_i - k_1 \sigma_i, p_i + k_2 \sigma_i], (p_i + k_2 \sigma_i, 1)\}$$

よって、求めるべき事後確率は次のようになる。D は減量(De-escalate), S は用量維持(Stay), E は増量(Escalate)を示す。

$$q(D, i) = p(p_i - p_t > k_2 \sigma_i | \text{data})$$

$$q(S, i) = p(-k_1 \sigma_i \leq p_i - p_t \leq k_2 \sigma_i | \text{data})$$

$$q(E, i) = p(p_i - p_t < k_1 \sigma_i | \text{data})$$

p_i の事後確率が $(0, p_i - k_1 \sigma_i)$ の範囲で最も大きい場合、つまり $q(E, i)$ が最大の場合、 p_i が p_t より小さいため増量、逆に $(p_i + k_2 \sigma_i, 1)$ の範囲で最も大きい場合、つまり $q(D, i)$ が最大の場合、 p_i は p_t より大きく危険な用量領域にあると判断し減量、とする。中央の領域である $[p_i - k_1 \sigma_i, p_i + k_2 \sigma_i]$ の範囲で最も大きい場合、つまり $q(S, i)$ が最大の場合、用量 i は真の MTD に近く用量を変えるべきではないと判断し、同じ用量を維持する、といった決定方法を用いる。

2.2 Ji らの方法の拡張

腎毒性や神経毒性ではグレード 3 とグレード 4 で大きく症状および重症度が異なるにもかかわらず、それらを同等に扱っている問題点に対する検討として、有害事象発現に関して、グレード 2 以下・グレード 3・グレード 4 の三項分布を仮定した。得られる観察データのうち、用量 i で治療される患者 n_i 人、グレード 3 での毒性発現者を x_i 人、グレード 4

での毒性発現者を y_i 人, グレード 3 での毒性発現確率を p_3 , グレード 4 での毒性発現確率を p_4 とすると, 尤度関数は三項分布の密度関数に従い次のようになる.

$$l(p) \propto \prod_{i=1}^d p_3^{x_i} p_4^{y_i} (1-p_3-p_4)^{n_i-x_i-y_i}$$

本研究では, 事前分布は, 多項分布の共役事前分布である Dirichlet 分布を用いた. Dirichlet 分布は, 連続型の確率分布であり, ベータ分布を多変量に拡張した分布である. パラメータ \mathbf{p} が $\text{Dir}(\alpha_1, \dots, \alpha_k)$ に従うとき, Dirichlet 分布の密度関数は次式で表わされる.

$$p(\mathbf{p}) = \frac{\Gamma(\alpha_1 + \dots + \alpha_k)}{\Gamma(\alpha_1) \dots \Gamma(\alpha_k)} \mathbf{p}_1^{\alpha_1-1} \dots \mathbf{p}_k^{\alpha_k-1}$$

$$\mathbf{p}_1, \dots, \mathbf{p}_k \geq 0; \sum_{j=1}^k \mathbf{p}_j = 1$$

ここで, 本研究で求めたいパラメータを $\mathbf{p}=(p_3, p_4)$ とおくと, 事前分布である Dirichlet 分布は次のように書ける. $\alpha_1, \alpha_2, \alpha_3$ は Dirichlet 分布のパラメータである. 本研究では, 2.5 で述べるベイズリスクの積分計算をする際に α が 1.5 以上である必要があったため, $\alpha_1=1.5, \alpha_2=1.5, \alpha_3=1.5$ を用いた.

$$p(\mathbf{p}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} p_3^{\alpha_1-1} p_4^{\alpha_2-1} (1-p_3-p_4)^{\alpha_3-1}$$

ここで, Dirichlet 分布は三項分布の共役事前分布であるため, 事後分布は次のように書ける.

$$p(\mathbf{p} | \text{data}) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} p_3^{\alpha_1-1+x} p_4^{\alpha_2-1+y} (1-p_3-p_4)^{\alpha_3-1+n-(x+y)}$$

2.3 損失関数

2.3.1 用いた損失関数

Ji らの方法の拡張点の 2 点目として, グレード 3 とグレード 4 の毒性発現確率に対する損失を仮定した. p_3, p_4 が目標毒性発現確率 p_t から遠ざかるほど重いリスクを想定し, グレード 3 よりもグレード 4 の方に大きい損失を設定した. 損失関数は様々考えられるが, $p_t = 1/3$ としたとき, 上に述べた条件を満たし, $p_t = p_3 + p_4 = 1/3$ の直線付近で損失が小さくなるような関数として, 二次元正規分布のような形状の曲面の関数を想定し, 本研究では次の関数を用いた.

$$\exp\left(2 \times (p_3 - 1/4)^2 + 3.2 \times (p_4 - 1/12)^2\right)$$

2.3.2 損失関数の提案

2.3.1 で用いた損失関数に加え, p_3, p_4 の分散と共分散を考慮した 2 変量正規分布の形の損失関数を提案する. p_3, p_4 の分散を σ_x^2, σ_y^2 , 共分散を ρ_{xy} とする.

2 変量正規分布の密度関数は次式で表わされる.

$$f(p_3, p_4) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho_{xy}^2}} \exp \left[-\frac{1}{2(1-\rho_{xy}^2)} \left\{ \frac{(p_3 - \mu_x)^2}{\sigma_x^2} + \frac{(p_4 - \mu_y)^2}{\sigma_y^2} - \frac{2\rho_{xy}(p_3 - \mu_x)(p_4 - \mu_y)}{\sigma_x\sigma_y} \right\} \right]$$

また, パラメータ $P = (p_3, p_4)$ の事後分布は $Dir(\alpha_1, \alpha_2, \alpha_3)$ であり,

$\alpha_0 = \alpha_1 + \alpha_2 + \alpha_3$ とおくと, それぞれの平均, 分散, 共分散は次式で表わされる.

$$E(p_3) = \frac{\alpha_1}{\alpha_0}, \quad E(p_4) = \frac{\alpha_2}{\alpha_0}$$

$$Var(p_3) = \frac{\alpha_1(\alpha_0 - \alpha_1)}{\alpha_0^2(\alpha_0 + 1)}, \quad Var(p_4) = \frac{\alpha_2(\alpha_0 - \alpha_2)}{\alpha_0^2(\alpha_0 + 1)}$$

$$Cov(p_3, p_4) = -\frac{\alpha_1\alpha_2}{\alpha_0^2(\alpha_0 + 1)}$$

これらを 2 変量正規分布の式に代入し, さらにグレード 3 とグレード 4 の重みを考慮し, 指数部内を次のようにした損失関数を提案する.

$$\frac{1}{2(1-\rho_{xy}^2)} \left\{ \frac{3(x-1/8)^2}{\sigma_x^2} + \frac{6(y-1/12)^2}{\sigma_y^2} - \frac{2\rho_{xy}(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} \right\}$$

2.4 用量増減決定方法

本研究では, 用量増減の決定にベイズリスクを利用した. Ji らの方法と同様, 毒性発現確率(0,1)を 3 つの区間に分け, 本研究では, Ji らの方法と同様,

$\{(0, p_i - k_1\sigma_i), [p_i - k_1\sigma_i, p_i + k_2\sigma_i], (p_i + k_2\sigma_i, 1)\}$ の 3 つの区間を用いた. 定数である k_1, k_2 の値によってデザインは変わる. 例えば $k_2 > k_1$ の場合は, 増量よりも減量する方を好むよ

うな保守的なデザインになり，逆に $k_1 > k_2$ の場合は増量を好むようなデザインになる．本研究では，危険な用量の方へ上がりすぎないように $k_2 > k_1$ とし， $k_1 = 1, k_2 = 1.5$ と設定した．そして，求めたいベイズリスクを上記の区間に適応させ， $(p_i + k_2\sigma_i, 1)$ でのベイズリスクを R_1 ， $[p_i - k_1\sigma_i, p_i + k_2\sigma_i]$ でのベイズリスクを R_2 ， $(0, p_i - k_1\sigma_i)$ でのベイズリスクを R_3 とおくと，求めるべきベイズリスクは次のように書ける．ここで， $L(p_3, p_4)$ を損失関数とする．

$$R_1 = \iint_{p_i + k_2\sigma_i < p_3 + p_4 < 1} L(p_3, p_4) \pi(p_3, p_4 | data) d p_3 d p_4$$

$$R_2 = \iint_{p_i - k_1\sigma_i < p_3 + p_4 < p_i + k_2\sigma_i} L(p_3, p_4) \pi(p_3, p_4 | data) d p_3 d p_4$$

$$R_3 = \iint_{0 < p_3 + p_4 < p_i - k_1\sigma_i} L(p_3, p_4) \pi(p_3, p_4 | data) d p_3 d p_4$$

用量増減の決定方法としては， R_1 が最も大きくなった場合は毒性発現確率が p_i より大きい範囲であるため，毒性が発現する危険な用量域であるとして減量， R_2 が最も大きくなった場合は p_i 付近であるため，用量を変更しない， R_3 が最も大きくなった場合は p_i にまだ達していないため，まだ用量を上げることができるとして増量，と判断した．

2.5 試験中止基準

次の場合には，試験中止とした．

- 1) 症例数が予定していた登録患者数に達した場合
- 2) p_i が次の条件を満たす場合

$$\tau_i = 1\{p(p_i > p_t | data) > \xi\}$$

ここで $1\{\}$ は指示関数 τ_i ， $\xi \in (0, 1)$ はカットオフ値とし，本研究では $\xi = 0.95$ とした． ξ が大きい値でかつ $\tau_i = 1$ のときは用量 i は毒性発現の可能性が極めて高く，用量 i へ増量すべきでないことを示す．

2.6 MTD の選択

用量が上がるにつれ毒性も上がるように設定するために，試験がすべて終了した後，MTD を選択する際に，isotonic 回帰を用いた．まず Dirichlet 分布の状況下で事後平均 \hat{p}_i を計算し， \hat{p}_i に PAVA (pooled adjacent violators algorithm: isotonic 回帰の 1 つ) を用いること

で、得られる \hat{p}_i^* は用量が上がるにつれ大きくなる。すなわち、 $j > i$ の条件下で、必ず $\hat{p}_j^* > \hat{p}_i^*$ となる。次に、MTD 選択の際のルールを以下に示す。

1) Ji らの方法と同様、用量増減はその時点での用量から 1 用量のみ増減を許すこととした。つまりある時点での用量を i ($i \in \{1, \dots, d\}$) とすると、その時点のコホートの毒性発現の有無を観察した後、次のコホートの用量を $i-1$, i , $i+1$ の 3 用量の中から選択する。例外として、 $i=1$ のときは次の用量は、 i もしくは $i+1$, $i=d$ のときは次の用量は $i-1$ もしくは i となる。

2) 試験終了時、用いた用量の中で $|\hat{p}_i^* - p_i|$ が最も小さくなり、かつ $\tau_i = 0$ である推定 MTD を選択する。もし 2 用量以上の $|\hat{p}_i^* - p_i|$ 同じ値となった場合は次のルールを適応した。 p^* を同じ値となった用量での \hat{p}_i^* とする。

a) $p^* > p_i$ のとき $|\hat{p}_i^* - p_i|$ が同じ値となった用量の中で最も高い用量を選択する。

b) $p^* < p_i$ のとき $|\hat{p}_i^* - p_i|$ が同じ値となった用量の中で最も低い用量を選択する。

2.7 提案法を用いた際の用量増減決定の結果

患者数が 6, 9 人の場合で提案法と Ji らの方法を用いた際の用量増減の結果を表 1, 2 に示す。各表の提案法の Decision で太字のところが、Ji らの方法と異なった結果になった決定である。このように、提案法の利点は、 p_i , n_i , x_i , y_i の値を与えれば、3 区間の事後確率 $q(D, i)$, $q(S, i)$, $q(E, i)$ をあらかじめ計算し、比較することができる点である。

$p_i = 0.3$ を用い、D: 用量 $i-1$ へ減量, S: 用量変更なし, E: 用量 $i+1$ へ増量, U: 毒性が高く危険な用量と判断し、用量 i は試験から除外すべき、を示す。

表 1. 用量 i での患者数と毒性発現者数に基づいた提案法の用量決定 (患者数 6 人の場合)

患者数	毒性発現数	グレード 3 毒性発現数	グレード 4 毒性発現数	提案法の Decision	Ji らの方法
6	0	0	0	増量	増量
	1	0	1	変更なし	変更なし
		1	0	増量	
	2	0	2	減量	減量
		1	1	減量	

患者数	毒性発現数	グレード3 毒性発現数	グレード4 毒性発現数	提案法の Decision	Ji らの方法
6	2	2	0	減量	減量
	3	0	3	減量	減量
		1	2	減量	
		2	1	減量	
		3	0	減量	

※以下，毒性発現者数 4 人以上の場合の Decision は，3 人の場合の Decision と同じ結果

表 2. 用量 i での患者数と毒性発現者数に基づいた提案法の用量決定（患者数 9 人の場合）

患者数	毒性発現数	グレード3 毒性発現数	グレード4 毒性発現数	提案法の Decision	Ji らの方法
9	0	0	0	増量	増量
	1	0	1	増量	増量
		1	0	増量	
	2	0	2	変更なし	変更なし
		1	1	変更なし	
		2	0	変更なし	
	3	0	3	変更なし	変更なし
		1	2	変更なし	
		2	1	変更なし	
		3	0	変更なし	
	4	0	4	減量	変更なし
		1	3	変更なし	
		2	2	変更なし	
		3	1	変更なし	
		4	0	変更なし	
	5	0	1	減量	減量
		1	4	減量	減量
		2	3	減量	
		3	2	減量	
		4	1	減量	
		5	0	減量	

※以下，毒性発現者数 6 人以上の場合の Decision は 5 人の場合の Decision と同じ結果

これらのスプレッドシート作成のために必要なものを表 3 にまとめた。

表 3. スプレッドシート作成に必要なもの

Ji らの方法	p_i , k_1 , k_2 , 事前分布 (ベータ分布) のパラメータ
提案法	p_i , k_1 , k_2 , 損失関数, 事前分布 (Dirichlet 分布) のパラメータ ここでは $\alpha_1=0.005$, $\alpha_2=0.005$, $\alpha_3=0.005$

3. シミュレーション

提案法と Ji らの方法を比較するために、以下の 2 通りの設定条件のもとで、シミュレーションを行った。シミュレーション回数は 1000 回とした。Ji らの方法で得られた結果を利用し、代表的な用量探索デザインである 3+3 デザイン、CRM とも比較を行った。

設定条件

- 1) 用量 8 段階, 1 コホート 3 人, 最大症例数 30 例, 目標毒性発現確率 $p_i=25\%$, 真の毒性発現確率は 6 通り (Scenario1~6) 設定した。真の毒性発現確率は結果の表の各 Scenario の 1 行目に記載した。
- 2) 1) と同様の設定条件で, グレード 3 よりグレード 4 の毒性発現確率が高くなるように真の毒性発現確率の設定を変更した。

そして、試験終了時に正しい MTD を選択する割合, 各用量での被験者数, 1 試験に必要な平均総症例数をシミュレーションの評価項目として, 提案法の性能を評価した。

3.1 シミュレーション結果

提案法と Ji らの方法, 3+3 デザイン, CRM との比較のシミュレーション結果を表 4, 表 5 に示した。表 4 はシミュレーションの評価項目である, 正しい MTD を選択する割合, 表 5 は各用量で治療される被験者数と平均症例数を示したものである。表 4 から, 想定した Scenario1~6 において, Scenario2, 6 を除き, 提案法では目標毒性発現確率 25% 以下となる最大の用量で正しい MTD を選択する割合が高かった。Scenario1 では真の MTD は用量 2 である。また, 真の MTD を選択した割合は, 3+3 デザインよりは高く Ji らの方法と CRM よりも低くなったが, MTD 以上の危険な用量を選択する割合は, 他のどの方法よりも提案法では低くなった。この傾向は他の Scenario にもあてはまる結果となった。Scenario2 は低用量では毒性が徐々に増加していき, 高用量では急激に毒性が増加するように真の毒性発現確率を想定した場合である。提案法では真の MTD である用量 7 ではなく, 1 つ下のレベルである用量 6 を MTD として選択する割合が最も高い結果となった。また,

Scenario3 は、用量 2 では毒性が低く、用量 3 で急に毒性が高くなる場合を想定したものである。Ji らの方法、3+3 デザイン共により結果となったが、提案法の性能は極めて高かった。Ji らの方法、提案法で他の方法よりも性能がよかったのは、中止基準を用いたためと考えられる。また Scenario4 は最初の用量で毒性発現確率が高く、ただちに試験を中止すべきであるという危険な状態を想定した場合であるが、提案法では過度の毒性のためにいかなる用量も MTD として選択しなかった割合が極めて高く、中止基準に基づき危険な用量で治療される患者数を除外するよう正しく MTD を推定できているといえる。Scenario5, 6 では毒性発現確率が徐々に上がっていく場合であるが、Scenario5 では、真の MTD は用量 2 で、それを選択する割合は提案法では Ji らの方法よりも高かった。用量 3 以上を選択した患者の割合もほとんどなく、Ji らの方法より精度よく MTD を推定していると考えられる。一方、Scenario6 では、正しい MTD は用量 3 と考えられるが、提案法では 1 つ下のレベルである用量 2 を選択する割合が最も高かった。このことから、Scenario2 と Scenario6 では MTD を低めに選択する傾向があったといえる。その原因が、今回用いた損失関数によるものか、方法論自体によるものか、検討する必要がある。

各用量で治療される被験者数は、ほとんどの場合で提案法の方が Ji らの方法よりも少ない人数となった。また 1 試験に必要な平均症例数は Ji らの方法では Scenario4 を除き最大の 30 例であったが、提案法ではそれに比べ少ない症例数となった。特に提案法では Ji らの方法と比べ、MTD よりも高い用量で治療される患者数が少ない傾向にあった。

また、表 6, 表 7 は、提案法を用いて同じ評価項目で、グレード 3 よりグレード 4 の毒性発現確率が高くなるように真の毒性発現確率を変更してシミュレーションを行った結果である。正しい MTD を選択する割合は、グレード 3 の発現確率の方が高い場合とあまり大きな違いは見られなかったが、Scenario2,6 ではこの場合の方がそれぞれ真の MTD と考えられる用量 7, 用量 3 での選択割合が高くなった。各用量で治療される被験者数も、ほとんど大差が見られなかったが、Scenario4 では、被験者数がより少なくなった。これはグレード 4 の方に重みをつけた損失関数の影響と考えられる。

提案法とJiらの方法, 3+3デザイン, CRMとの比較のシミュレーション結果

表 4. 正しい MTD を選択する割合(%)

真: グレード 3 と 4 を合わせた真の毒性発現確率, None: 過度の毒性のためにいかなる用量も MTD として選択しなかった割合

Scenario	Dose	1	2	3	4	5	6	7	8	None
1	真	5	25	50	60	70	80	90	95	
	(Grade3/Grade4)	(4/1)	(20/5)	(40/10)	(45/15)	(50/20)	(55/25)	(60/30)	(63/32)	
	提案法	21	74	2	0	0	0	0	0	3
	Ji	13	79	8	0	0	0	0	0	0
	3+3	24	58	16	2	0	0	0	0	0
	CRM	6	83	11	0	0	0	0	0	0
2	真	1	2	3	4	5	6	25	50	
	(Grade3/Grade4)	(0.8/0.2)	(1.4/0.6)	(2/1)	(2.8/1.2)	(3.5/1.5)	(4.2/1.8)	(20/5)	(40/10)	
	提案法	0	0	2	5	15	42	36	0	0
	Ji	0	0	0	0	2	22	62	14	0
	3+3	0	0	0	2	3	21	46	8	0
	CRM	0	0	0	1	1	5	22	50	
3	真	1	5	50	60	70	80	90	95	
	(Grade3/Grade4)	(0.8/0.2)	(4/1)	(40/10)	(45/15)	(50/20)	(55/25)	(60/30)	(63/32)	
	提案法	0	90	10	0	0	0	0	0	0
	Ji	0	79	21	0	0	0	0	0	0
	3+3	0	70	28	2	0	0	0	0	0
	CRM	0	49	51	0	0	0	0	0	0

Scenario	Dose	1	2	3	4	5	6	7	8	None
4	真 (Grade3/Grade4)	40 (35/5)	50 (40/10)	60 (45/15)	70 (50/20)	80 (55/25)	90 (60/30)	95 (63/32)	99 (65/34)	
	提案法	8	7	0	0	0	0	0	0	85
	Ji	31	2	0	0	0	0	0	0	67
	3+3	38	9	1	0	0	0	0	0	52
	CRM	47	2	0	0	0	0	0	0	51
5	真 (Grade3/Grade4)	15 (13/2)	25 (20/5)	35 (27/8)	45 (35/10)	55 (43/12)	65 (50/15)	75 (55/20)	85 (60/25)	
	提案法	12	70	4	0	0	0	0	0	14
	Ji	31	41	21	7	0	0	0	0	0
	3+3	29	37	20	7	1	0	0	0	8
	CRM	36	47	14	2	0	0	0	0	0
6	真 (Grade3/Grade4)	5 (4/1)	15 (13/2)	25 (20/5)	35 (27/8)	45 (35/10)	55 (43/12)	65 (50/15)	75 (55/20)	
	提案法	5	60	33	0	0	0	0	0	2
	Ji	2	24	42	24	7	0	0	0	0
	3+3	9	28	34	22	5	0	0	0	0
	CRM	4	37	45	12	2	0	0	0	0

提案法と Ji の方法, 3+3 デザイン, CRM との比較のシミュレーション結果

表 5. 各用量で治療される被験者数 (人)

Scenario	Dose	1	2	3	4	5	6	7	8	平均症例数
1	真	5	25	50	60	70	80	90	95	
	((Grade3/Grade4)	(4/1)	(20/5)	(40/10)	(45/15)	(50/20)	(55/25)	(60/30)	(63/32)	
	提案法	3.8	22.4	3.1	0	0	0	0	0	29
	Ji	7.7	16.1	5.8	0.5	0	0	0	0	30
	3+3	4.0	5.0	2.6	0.4	0	0	0	0	12
2	CRM	5.7	18.6	4.9	1.0	0	0	0	0	30
	真	1	2	3	4	5	6	25	50	
	((Grade3/Grade4)	(0.8/0.2)	(1.4/0.6)	(2/1)	(2.8/1.2)	(3.5/1.5)	(4.2/1.8)	(20/5)	(40/10)	
	提案法	0.2	3.2	4.6	4.9	4.8	6.5	4.8	1.0	30
	Ji	3.2	3.2	3.4	6.5	6.7	4.5	5.9	2.6	30
3	3+3	3.1	3.2	3.3	3.4	3.3	3.7	4.5	2.2	27
	CRM	3.1	3.4	3.3	3.7	3.6	4.4	5.1	3.4	30
	真	1	5	50	60	70	80	90	95	
	((Grade3/Grade4)	(0.8/0.2)	(4/1)	(40/10)	(45/15)	(50/20)	(55/25)	(60/30)	(63/32)	
	提案法	0.1	19.3	10.6	0	0	0	0	0	30
	Ji	5.5	13.2	10.2	1.0	0	0	0	0	30
	3+3	3.1	5.2	4.4	0.7	0.1	0	0	0	13
	CRM	3.1	13.0	12.0	1.8	0	0	0	0	30

Scenario	Dose	1	2	3	4	5	6	7	8	平均症例数
4	真	40	50	60	70	80	90	95	99	
	(Grade3/Grade4)	(35/5)	(40/10)	(45/15)	(50/20)	(55/25)	(60/30)	(63/32)	(65/34)	
	提案法	5.0	8.6	0	0	0	0	0	0	14
	Ji	16.8	1.8	0.2	0	0	0	0	0	19
	3+3	4.7	0.5	0.6	0.7	0	0	0	0	6
	CRM	20.2	2.5	0.2	0	0	0	0	0	23
5	真	15	25	35	45	55	65	75	85	
	(Grade3/Grade4)	(13/2)	(20/5)	(27/8)	(35/10)	(43/12)	(50/15)	(55/20)	(60/25)	
	提案法	2.9	22.3	2.0	0	0	0	0	0	27
	Ji	12.4	9.5	5.5	1.9	0.3	0	0	0	30
	3+3	4.4	3.9	2.4	0.9	0.2	0	0	0	12
	CRM	13.8	11.4	3.6	0.9	0.2	0	0	0	30
6	真	5	15	25	35	45	55	65	75	
	(Grade3/Grade4)	(4/1)	(13/2)	(20/5)	(27/8)	(35/10)	(43/12)	(50/15)	(55/20)	
	提案法	1.3	21.0	7.3	0	0	0	0	0	30
	Ji	5.1	8.2	9.2	5.7	1.6	0.3	0	0	30
	3+3	3.6	4.3	3.8	2.3	0.8	0.2	0	0	15
	CRM	5.5	11.5	8.9	3.4	0.7	0.1	0	0	30

提案法のシミュレーション結果(Grade4>Grade3の場合)

表 6. 正しい MTD を選択する割合(%)

真: グレード3と4を合わせた真の毒性発現確率, None: 過度の毒性のためにいかなる用量も MTD として選択しなかった割合

Scenario	Dose	1	2	3	4	5	6	7	8	None
1	真	5	25	50	60	70	80	90	95	
	(Grade3/Grade4)	(1/4)	(10/15)	(20/30)	(25/35)	(30/40)	(35/45)	(40/50)	(43/52)	
	提案法	25	63	2	0	0	0	0	0	10
2	真	1	2	3	4	5	6	25	50	
	(Grade3/Grade4)	(0.6/0.4)	(0.9/1.1)	(1.5/1.5)	(1.8/2.2)	(2/3)	(2.5/3.5)	(7/18)	(20/30)	
	提案法	0	0	2	6	12	40	40	0	0
3	真	1	5	50	60	70	80	90	95	
	(Grade3/Grade4)	(0.7/0.3)	(1/4)	(25/25)	(27/33)	(30/40)	(35/45)	(40/50)	(43/52)	
	提案法	0	86	13	0	0	0	0	0	1
4	真	40	50	60	70	80	90	95	99	
	(Grade3/Grade4)	(15/25)	(20/30)	(25/35)	(30/40)	(35/45)	(40/50)	(43/52)	(45/54)	
	提案法	8	2	0	0	0	0	0	0	90
5	真	15	25	35	45	55	65	75	85	
	(Grade3/Grade4)	(5/10)	(10/15)	(15/20)	(20/25)	(25/30)	(30/35)	(35/40)	(40/45)	
	提案法	20	41	10	0	0	0	0	0	29
6	真	5	15	25	35	45	55	65	75	
	(Grade3/Grade4)	(2/3)	(5/10)	(10/15)	(15/20)	(20/25)	(25/30)	(30/35)	(35/40)	
	提案法	6	42	40	5	0	0	0	0	7

提案法のシミュレーション結果(Grade4>Grade3 の場合)

表 7. 各用量で治療される被験者数 (人)

Scenario	Dose	1	2	3	4	5	6	7	8	平均症例数
1	真	5	25	50	60	70	80	90	95	
	(Grade3/Grade4)	(1/4)	(10/15)	(20/30)	(25/35)	(30/40)	(35/45)	(40/50)	(43/52)	
2	提案法	4.7	15.1	7.5	0.5	0	0	0	0	28
	真	1	2	3	4	5	6	25	50	
	(Grade3/Grade4)	(0.6/0.4)	(0.9/1.1)	(1.5/1.5)	(1.8/2.2)	(2/3)	(2.5/3.5)	(7/18)	(20/30)	
3	提案法	0.2	3.3	4.4	4.9	4.6	6.5	4.9	1.1	30
	真	1	5	50	60	70	80	90	95	
	(Grade3/Grade4)	(0.7/0.3)	(1/4)	(25/25)	(27/33)	(30/40)	(35/45)	(40/50)	(43/52)	
4	提案法	0.6	15.5	13.2	0.6	0	0	0	0	30
	真	40	50	60	70	80	90	95	99	
	(Grade3/Grade4)	(15/25)	(20/30)	(25/35)	(30/40)	(35/45)	(40/50)	(43/52)	(45/54)	
5	提案法	4.6	4.7	1.3	0	0	0	0	0	11
	真	15	25	35	45	55	65	75	85	
	(Grade3/Grade4)	(13/2)	(20/5)	(27/8)	(35/10)	(43/12)	(50/15)	(55/20)	(60/25)	
6	提案法	4.4	11.0	7.2	1.3	0.1	0	0	0	24
	真	5	15	25	35	45	55	65	75	
	(Grade3/Grade4)	(4/1)	(13/2)	(20/5)	(27/8)	(35/10)	(43/12)	(50/15)	(55/20)	
	提案法	2.0	11.0	11.7	3.0	0.6	0	0	0	18

5. 考察

本研究では、グレード3とグレード4における毒性の重症度の違いに着目し、ベイズ流デザインを用いた新たな用量探索方法を提案した。提案法を用いることにより、患者ごとの毒性の重症度を考慮した用量探索が可能となるといえる。本研究は、先行研究である Ji らの方法を拡張したものであるが、提案法 Ji らの方法の特徴である、方法自体のシンプルさは拡張したことでも失われておらず、より臨床医にとって患者の症状に合わせて用いることができるデザインであるといえる。毒性の重症度を考慮した既存の提案法([3])はとても複雑なデザインである一方、本研究の提案法は同じ程度の精度を保ちつつ、よりシンプルなデザインであるといえる。実際の医療現場では、臨床医がどの用量探索方法を用いるかどうかを決めるため、方法自体の簡易さは重要である。

もう1つの利点は、第2章に述べたように、必要項目を設定することで、患者数と毒性発現者数に基づいた用量決定のスプレッドシートをあらかじめ作成できることである。また試験をモニタリングしている際においてもこのスプレッドシートを用いて用量決定を迅速に行うことができる。

提案法を用いて用量決定を行い、シミュレーションをするのに必要なのは目標毒性発現確率 p_t 、用量のレベル数、最大症例数、真の毒性発現確率、損失関数である。シミュレーションにより、 p_t 付近では小さく、 p_t から遠ざかるほど大きな、また重症度の高いグレード4の方に重い損失を事後確率にかけることで、Ji らの方法と比べてグレード4の平均毒性発現者数を少なくすることができるといえる。多くの新規治療薬が開発される一方で、臨床試験に参加する症例数には限りがあり、がん第I相試験の対象患者は、有効な標準治療が存在しない進行期の患者である。従って、効果が低いあるいは安全性に問題があることが分かれば早期に試験を中止する必要がある。その点で、提案法では危険な用量の方へ割り当てられる患者数を減らすことができ、より安全な試験を行えるといえる。

提案法の特徴である損失関数であるが、どのような損失関数を設定するかが結果に影響を及ぼす大きな一因でもあったと考えられる。今回用いた損失関数の曲面が、最も適した関数とは限らず、パラメータを変えた曲面や別の形状の曲面など、さらに検討の余地があるといえる。提案した損失関数についてもパラメータを調整してシミュレーションで検証したい。

中止基準で用いた ξ は本研究では、 $\xi=0.95$ を用いたが、より小さい値を用いることで用量をより除外しやすくなる。よって保守的なデザインとなる。実際、M.D. Anderson Cancer Center では $\xi=0.70$ を用いているようである。シミュレーションの結果から、Scenario の真の毒性発現確率に応じて、危険な用量域に入らないように十分に用量を除外できていたのではないかと考えられる。また、Dirichlet 分布のパラメータも結果に影響を与える要因であると考えられる。今回は1通りしか検討しなかったが、更に検討してみる必要がある。

どの Scenario においても 3+3 デザインよりは精度がよく, Ji らの方法と比べた場合には真の毒性発現確率によって精度の良し悪しが異なる結果となったが, 正しい MTD を推定するのに用いることができる十分な結果が得られたと考えられる. 調整できるパラメータを変えることで, より精度のよい結果が得られると示唆される.

最後に, 異なる二つの有害事象が同じグレードであっても症状の重症度が異なる場合においても, この提案法を用いることにより重症度が高い方の有害事象の発現を抑えることができると考えられる. 今後の課題として, 肝毒性や腎毒性のような発現率は低いが重症度が高いような毒性と, 骨髄抑制や嘔吐のように高頻度で発現するが重症度は低い毒性に対しそれぞれ重症度を反映した損失関数を考え, 発現する毒性に応じて損失関数を使い分けして用量を決定できるようなシミュレーションを行いたいと思う.

6. 参考文献

- [1] Crowley J. (2005). Handbook of Statistics in Clinical Oncology 2nd edition. New York: Marcel Dekker Inc.
- [2] O'Quigley J, Pepe M, Fisher L. Continual reassessment method: a practical design for phase I clinical trials in cancer. *Biometrics* 1990; 46: 33-48
- [3] Yuan Ji, Yisheng Li, B.Nebiyu Bekele, (2007). Dose-finding in phase I clinical trials based on toxicity probability intervals. *Clinical Trials* 4:235-244
- [4] B.Nebiyu Bekele, Peter F.Thall. Dose-Finding Based on Multiple Toxicities in a soft Tissue Sarcoma Trial. *Journal of the American Statistical association*. March 2004, vol.99,No.465, Applications and Case Studies
- [5] Z.Yuan, R.Chappell, H.Bailey. The Continual Reassessment Method for Multiple Toxicity Grades: A Bayesian Quasi-Likelihood Approach. *Biometrics* 63,173-179, March 2007
- [6] B.Nebiyu Bekele, Yisheng Li, Yuan Ji. Risk-Group-Specific Dose Finding Based on an Average Toxicity Score. *Biometrics*, April 2009
- [7] B.Neuenschwander, M.Branson, T.Gsponer. Critical aspects of the Bayesian approach to phase I cancer trials. *Statistics In Medicine*,2008;27:2420-2439
- [8] D.M.Potter. Phase I Studies Of Chemotherapeutic Agents In Cancer Patients:A Review Of The Designs. *Journal of Biopharmaceutical Statistics*,16:579-604,206
- [9] Andrew Gelman, John B.Carlin, Hal S.Stern, and Donald B.Rubin, Bayesian Data Analysis, Chapman&Hall/CRC, (2000)
- [10] Common Terminology Criteria for Adverse Events v3.0 (CTCAE), NCI, (2006)
- [11] 渡部洋, ベイズ統計学入門, 福村出版, (1999)
- [12] 福田治彦, 新美三由紀, 石塚直樹訳, 米国 SWOG に学ぶがん臨床試験の実践—臨床医と統計家の協調をめざして, 医学書院, (2004)